Sequence-independent RNA cleavages generate the primers for plus strand DNA synthesis in hepatitis B viruses: implications for other reverse transcribing elements

Daniel D.Loeb¹, Russell C.Hirsch¹ and Don Ganem^{1,2}

¹Departments of Microbiology and Immunology and ²Medicine, Howard Hughes Medical Institute, University of California Medical Center, San Francisco, CA 94143, USA

Communicated by H.Schaller

Reverse transcription of RNA into duplex DNA requires accurate initiation of both minus and plus strand DNA synthesis; this in turn requires the generation of specific primer molecules. We have examined plus strand primer generation in the hepatitis B viruses, small DNA viruses that replicate via reverse transcription. The plus strand primer in these viruses is a short capped RNA derived from the 5' end of the RNA template by cleavage at a specific set of sites. To elucidate the cleavage mechanism we constructed a series of viral mutants bearing alterations in and around the cleavage sites. Our results reveal that the cleavage reaction is sequence-independent and indicate that the cleavage sites are positioned by measurement of the distance from the 5' end of the RNA. Comparison of these findings with what is known about RNase H-mediated primer generation in retroviruses and other retroid elements suggests that, despite many divergent features, some common molecular features are preserved.

Key words: hepadnaviruses/plus strand primer/reverse transcription/RNase H

Introduction

Reverse transcription, a process once thought to be restricted to vertebrate retroviruses (Baltimore, 1970; Temin and Mizutami, 1970), is now known to occur in many different classes of eukaryotic genetic elements, including several families of plant and animal viruses, many transposons of lower and higher eukaryotes, some fungal mitochondrial plasmids and bacterial minus strand DNA (Summers and Mason, 1982; Pfeifer and Hohn, 1983; Boeke *et al.*, 1985; Kuiper and Lambowitz, 1988; Inouye *et al.*, 1989). In addition, 5% or more of the vertebrate genome is composed of repeated sequences believed to have been derived by reverse transcription (Deininger, 1989; Hutchinson *et al.*, 1989). An understanding of the reverse transcription reaction is therefore of considerable importance. Although the mechanics of retroviral reverse transcription are understood in some detail (cf. Varmus and Brown, 1989), many of the other elements employing RNA-directed DNA synthesis appear to employ quite different strategies for generating duplex DNA from RNA. Elucidation of the details of these strategies should identify both common themes essential to all reverse transcription reactions and variations that give the products their rich diversity of structure and function.

All known reverse transcriptases (RTs) share distinct, conserved amino acid homologies associated with DNA polymerase (Toh *et al.*, 1983) and RNaseH activities (Johnson *et al.*, 1986; Tanese and Goff, 1988; Radziwill *et al.*, 1990). The latter is required to remove the RNA template following first (minus) strand DNA synthesis, thereby allowing this strand to serve as template for the synthesis of the second (plus) DNA strand. RT also shares with other polymerases the inability to initiate DNA synthesis *de novo*, requiring a primer to which to add the first deoxy-nucleotide. Since the accurate initiation of DNA synthesis is critical for formation of a properly structured DNA product, the priming steps in minus and plus strand synthesis represent key events in the reaction. In retroviruses, plus strand synthesis is primed by a short oligoribonucleotide derived from the RNA template by the action of RNaseH (Finston and Champoux, 1984; Resnick *et al.*, 1984; Smith *et al.*, 1984), and some progress has been made in understanding the cleavage mechanism that generates this primer (Oyama *et al.*, 1989; Furfine and Reardon, 1991). In this paper we examine the mechanism of plus strand primer generation in another class of reverse transcribing elements, the hepatitis B virus (hepadnaviruses), whose reverse transcription pathway is quite divergent from that of retroviruses (see Ganem and Varmus, 1987). Our results suggest that despite many differences in the mechanics of priming, some fundamental features of the reaction are conserved.

Results

Experimental design

In this study we have used the duck hepatitis B virus (DHBV) to study the events of plus strand primer generation. Figure 1 illustrates the current model for DHBV DNA synthesis. Mature DHBV DNA is a relaxed circular duplex of 3 kb. The 5' ends of the minus and plus DNA strands map to direct repeats (DRs) of 12 nucleotides (Molnar-Kimber *et al.*, 1984) known as DR1 and DR2, respectively; these repeats are involved in priming the synthesis of each DNA strand. The RNA template for reverse transcription, the RNA pregenome (Buscher *et al.*, 1985; Enders *et al.*, 1985), is terminally redundant and carries two copies of DR1, but only a single copy of DR2 (Figure 1, row i). Following its encapsidation into a subviral core particle (Hirsch *et al.*, 1990), minus strand synthesis initiates within the 3' DR1 sequence (Seeger and Maragos, 1990), in a reaction believed to be primed by a terminal protein (Bartenschlager and Schaller, 1988; Figure 1, row ii). Synthesis of minus strand DNA is carried out by the RT encoded by the viral P gene (Summers and Mason, 1982; Radziwill *et al.*, 1990; Chang *et al.*, 1990); during elongation, the RNaseH activity of the P protein degrades the recently copied RNA (Figure 1, row iii). Full-length minus DNA has a short terminal redundancy of 8–10 nucleotides, termed 'r' (Lien *et al.*, 1987). Upon

Fig. 1. A model of the hepadnaviral replication pathway. Dashed line (row i) represents pregenomic RNA, which is capped and terminally redundant (R) by ~270 nucleotides. Open rectangular boxes labeled 1 and 2 represent DR1 and DR2 respectively. After selective packaging of this RNA molecule into a cytoplasmic subviral core particle, reverse transcription commences. Minus strand DNA synthesis (solid line, row ii) is initiated within the 3' copy of DR1 and is primed by a terminal protein encoded by the P gene product. Minus strand DNA synthesis proceeds with the concomitant degradation of the RNA template by the viral RNase H (row iii). Upon completion of minus strand synthesis (row iv) the primer for plus strand DNA synthesis is generated comprising a short capped oligoribonucleotide derived from the 5' end of pregenomic RNA. The short terminal redundancy in minus DNA is indicated by 'r'. This primer is either translocated to the downstream DR2 site on minus DNA (row v, left) or extended *in situ* (row v, right). The latter pathway results in a duplex linear genome that is non-infectious. If translocated and elongated, circularization of the genome occurs during plus strand DNA synthesis resulting in the relaxed circular form of the genome found in mature infectious virions (row vi). Underneath the two mature DNA forms is indicated the position of the two oligonucleotides, A and B, used in the primer extension analyses described in this paper. Oligonucleotide A can be used to determine the 5' end of plus strand DNA on both linear duplex and relaxed circular molecules, while oligonucleotide B will determine the 5' end of plus DNA on only relaxed circular molecules since it is ~3 kb from the 5' end of plus DNA on linear duplex molecules.

completion of minus strand synthesis, the primers for plus strand synthesis are generated (Lien *et al.*, 1987). These are short (15–18 nucleotides) capped oligoribonucleotides cleaved from the 5' end of pregenomic RNA at four sites near the 3' border of DR1 (Figure 1, row iv). Since plus strand synthesis initiates at DR2, the plus strand primer must be translocated to this site, to which it can be base-paired due to the identity of the DRs (Figure 1, row v, left). After DNA synthesis proceeds to the 5' end of minus DNA, an intramolecular jump (presumably mediated by the homologous r sequences) allows circularization of the genome and completion of plus strand synthesis (Figure 1, row vi).

One of the most remarkable and distinctive steps in this pathway is that of plus strand priming. Here we examine this step in detail, with special attention to the following questions: what directs the cleavage enzyme to cleave at the 3' end of DR1? To define the cleavage mechanism we have constructed mutations in and around the cleavage sites. Mutant DHBV genomes bearing these lesions are transfected into LMH cells, an avian hepatoma line that is permissive for viral DNA synthesis (Condreay *et al.*, 1990). Several days later, cytoplasmic core particles containing viral replicative intermediates are isolated; viral DNA extracted

from these cores is then examined by Southern hybridization to determine the gross effects of the lesion on the levels of viral DNA production and to look for major effects on the forms of viral DNA produced. (Wild-type DHBV produces large quantities of the relaxed circular viral DNA and smaller amounts of duplex linear and nascent single stranded minus species schematically depicted in Figure 1). The sites of primer cleavage are determined by an indirect assay: after chemically removing the RNA primer from plus strand DNA, we used primer extension to determine the position of the 5' end of the DNA (Seeger and Maragos, 1989; Staprans *et al.*, 1991). That is, we infer the location of the sites of primer cleavage by mapping the sites of the RNA–DNA junctions. Mutations that inactivate cleavage would be expected to prevent plus strand DNA synthesis, while mutations that alter cleavage specificity should alter the 5' end of plus DNA. The two oligonucleotides (A and B) employed in these assays are depicted in Figure 1; except where specifically noted, the data presented below were derived using oligonucleotide A.

This indirect assay of priming sites used *in vivo* is necessary because no soluble *in vitro* system has been developed in which the complete hepadnaviral replication

Fig. 2. Sequence within DR1 does not direct cleavage specificity. (A) Sequence of the 5' end of pregenome in the 5/12 mutant. Boxed sequence indicates the position of DR1. Bold-faced nucleotides denoted by asterisk represent the substitution mutations within DR1 in the 5/12 mutant. (B) Southern blot analysis of wild type (lane 1) and the 5/12 mutant (lane 2) viral DNA from transfected LMH cells. On left is indicated the positions of relaxed circular (RC), duplex linear (L), and single stranded minus (SS) DNA. (C) Primer extension analysis to determine the 5' of plus DNA of the 5/12 mutant. Lanes 3 to 6, sequencing ladder of cloned wild type DHBV DNA generated with ³²P-end labeled oligonucleotide A; lane 7, primer extension on 5/12 viral DNA using ³²P-end labeled oligonucleotide A. Four arrows on the right indicates the 5' ends mapped, which are the same as found in wild type linear molecules (see Staprans *et al.*, 1991; Figure 7). The position of DR1 is indicated by black vertical bar on left. The positions of the determined cleavage sites are indicated by the arrows beneath the sequence in panel A.

reaction occurs. The viral replication machinery is localized within tightly organized core particles and does not accept exogenous primer/templates (Radziwill *et al.*, 1988), so it is difficult to assay directly for the cleavage reaction by presenting model substrates to these complexes. Clearly, only primers that are functional for elongation can be assayed in our *in vivo* assay; however, we have recently found that even mutants with highly abnormal cleavage sites support wild-type levels of plus strand DNA synthesis indicating that virtually all mutant primers can be examined in this way (Staprans *et al.*, 1991; see also below).

Sequences within DR1 do not direct the cleavage specificity

In earlier studies, we showed that mutations at the nucleotides directly flanking a major cleavage site did not prevent plus strand synthesis and therefore did not prevent primer cleavage (Staprans *et al.*, 1991). Rather, they allowed correct and efficient cleavage but inhibited primer translocation (indicating that these steps are genetically separable). These untranslocated primers were nonetheless efficiently elongated; as shown in Figure 1 (row v, right), the viral

DNA that results from extension of such primers is linear rather than circular. These studies suggest that the specificity of cleavage might not be encoded by the sequence directly at the cleavage site. Accordingly, we constructed additional mutations in other nearby regions to better define the cleavage determinants. An obvious potential signal for recognition by the cleavage machinery would be DR1 itself. To test this possibility we changed 5 of the 12 nucleotides within DR1 (Figure 2A). The mutations were introduced into the 5' copy of DR1, leaving a normal copy of DR1 in the 3' position so as not to disturb minus strand initiation and elongation.

LMH cells were transfected with this clone, and 3 days later cytoplasmic cores were harvested and their DNA examined by Southern blot analysis (Figure 2B). Although the overall level of DNA synthesis directed by this mutant was comparable to wild type, all duplex viral DNA was found as 3 kb linear species, with no detectable relaxed circular form. As described above, this indicates that functional plus strand primers were generated but were unable to be translocated and instead initiated plus strand DNA synthesis *in situ*. The 5' end(s) of the mutant plus strand DNA were then mapped by primer extension (Figure 2C, lane 7), using oligonucleotide A as a primer (see Figure 1 and Materials and methods). This showed that the RNA–DNA junctions map to four consecutive positions at the 3' end of DR1. These four positions are identical to those found in the linear duplex DNA from a wild type transfection (Staprans *et al.*, 1991), indicating that the efficiency and specificity of the cleavage reaction is unperturbed in the 5/12 mutant. This suggests that DR1 sequences play little if any role in determining the specificity of the cleavage reaction, a conclusion that is sustained by analysis of still more drastic mutations in this element (see below).

The specificity of cleavage is not determined by measurement from a determinant 3' of the cleavage site

Since mutations at and around the cleavage sites did not affect the cleavage reaction, it seemed unlikely that a conventional sequence-specific cleavage mechanism analogous to that used by class II restriction enzymes was involved. However, it remained possible that the enzyme functions by recognizing a specific sequence determinant then cleaving at a fixed distance from this sequence. This general model can be subdivided into two classes, according to whether the determinant is located 3' of the cleavage site or 5' of it. These models can be simply tested by placing either small insertions or deletions between the putative determinant and the cleavage site; such lesions should shift the cleavage sites in a predictable fashion if cleavage is dictated by measurement from such a determinant.

We first tested the model that the specificity of cleavage was being determined by measurement from a 3' determinant. To test this we introduced a three nucleotide insertion at the 3' boundary of DR1, just downstream of the wild type cleavage sites (Figure 3A). The sequence of the three inserted nucleotides was chosen to be the same as the three nucleotides immediately 3' of DR2, to circumvent any problem with base pairing of the 3' end of the RNA primer to the DR2 region if the primer were properly translocated. If the cleavage sites are being measured from a 3' determinant, then after cleavage the three inserted nucleotides will be

Fig. 3. Measurement from a 3' determinant does not specify cleavage. (A) Sequence of the 5' end of this +3 insertion mutant. Boxed sequence indicates DR1. The three asterisked bold-faced nucleotides at the 3' boundary of DR1 represent the insertion. Shaded circle denotes the arbitrary position of a hypothetical 3' measurement determinant. Arrowed lines above pregenome represent the two different primer extension products expected with oligonucleotide A or B with and without measurement from a 3' determinant. (B) Southern blot analysis of wild type (lane 1) and insertion mutant (lane 2) viral DNA from transfected LMH cells. On left is indicated the positions of relaxed circular (RC), duplex linear (L), and single stranded minus (SS) DNA. (C) Primer extension analysis with oligonucleotide A to map the 5' end of linear plus DNA in the mutant. Lane 3, mutant viral DNA with the arrows on the left indicating ends; lanes 4–7, sequencing ladder of cloned wild type DNA using end labeled oligonucleotide A. Black vertical bar on right indicates the position of DR1. Although use of this primer will also allow mapping of the 5' end of RC DNA (see Figure 1), only the region of the gel corresponding to the ends of linear species is shown. (D) Primer extension analysis with oligonucleotide B to determine the 5' end of plus DNA of the RC species of the mutant. Lane 8, mutant viral DNA with the 5' ends indicated by the arrows on the left; lanes 9–12, sequencing ladder of cloned wild type DNA generated with end labeled oligonucleotide B. Black vertical bar on right indicates the position of DR2. The positions of the mapped cleavages on the sequence are depicted by arrows in panel A.

included in the new RNA primer (see Figure 3A); that is, the cleavage site will move three nucleotides in the 3' direction. This will result in a plus strand DNA primer extension product that will be three bases shorter than wild type in the RC molecules and the same length as wild type in the duplex linear molecules. Southern blot analysis of viral DNA reveals a 1:1 ratio of RC to linear duplex molecules (Figure 3B). Determination of the 5' termini of these DNA molecules indicates that the predictions of the model are not met. Linear duplex molecules have 5' ends which are three nucleotides longer than wild type (Figure 3C) and relaxed circular molecules have 5' ends which are the same size as wild type (Figure 3D). This result indicates that the insertion had no effect on the cleavage specificity: the mutant DNA was cleaved at the wild-type sites. This argues strongly that the site of cleavage is not being determined by measurement from a determinant 3' of the cleavage site and further confirms that cleavage is largely independent of sequences immediately 3' of the cleavage site.

Insertions and deletions between the 5' end of pregenome and the wild type cleavage site move the position of the cleavage

To test the model that the cleavage sites are measured from a determinant 5' of these sites we introduced one, two (two different mutants), or four nucleotides into the 'r' region

immediately upstream of DR1 in the 5' region of pregenomic RNA. The sequences of these mutants are shown in Figure 4A. (In designing the +2b mutant and the +4 mutant the nucleotides inserted were derived from the sequence 5' of DR2 such that primers containing these insertions could still form a 12 base-pair hybrid with the DR2 region in minus DNA). If this model is correct then such lesions should move the cleavages in the 5' direction relative to the wild type site. LMH cells were transfected with the mutant DNAs and cytoplasmic cores containing progeny DNA examined by Southern blot analysis (Figure 4B). All of these insertion mutants are replication-competent, producing normal quantities of minus and plus strand DNA; the +2 and +4 insertion mutants produce disproportionate amounts of duplex linear DNA.

To map the positions of the cleavages we carried out primer extension analysis as before, with one exception. When mapping the 5' ends of plus DNA of these RC molecules there is an additional consideration: during the circularization of viral DNA, the nascent plus strand may or may not copy an insertion in 'r', depending on the exact mechanism of intramolecular circularization (i.e. the exact base at which the strand transfer occurs). Since at the onset of these studies this mechanism was not well understood, to circumvent these uncertainties we used oligonucleotide B (cf. Figure 1) to determine the 5' end of the RC molecules.

Fig. 4. The position of cleavage remains a constant distance from the 5' end of pregenome. (A) Sequence of the 5' end of pregenome of wild type (wt) and four mutants (+1, +2a, +2b, and +4). Boxed sequence is DR1. Position and sequence of insertions is indicated by asterisks over bold-face residues. (B) Southern blot analysis of viral DNA isolated from subviral core particles of transfected LMH cells. On left is indicated the positions of relaxed circular (RC), duplex linear (L), and single-stranded minus (SS) DNA. Lane 1, wild type viral DNA; lane 2, +1 insertion viral DNA; lane 3, +2a insertion viral DNA; lane 4, +2b insertion viral DNA; lane 5, +4 insertion viral DNA. (C) Primer extension analysis with oligonucleotide B to determine the 5' end of relaxed circular plus strand DNA. On left black bar indicates the position of DR2. Lanes 6–9, sequencing ladder of cloned wild type DHBV DNA generated with end labeled oligonucleotide B. Lanes 10–14, primer extension products of DNAs corresponding to the mutants indicated above each lane. Position of cleavages in each mutant is indicated in panel A with arrows underneath respective sequences.

This oligonucleotide anneals 5' of the 'r' region on plus DNA and allows direct determination of the 5' ends irrespective of events at r (see Figure 1 and legend). Primer extension analysis (Figure 4C) shows that the 5' ends of plus DNA are shifted two and four nucleotides upstream with the +2 and +4 mutants respectively, exactly as predicted by the 5' measurement model. The +1 mutant has plus strands that are two nucleotides longer than wild type; thus, though the cleavage in this mutant moved in the predicted direction, the extent of the movement is off by one base. The source of this discrepancy is not understood.

An additional prediction of this model is that deletions between the 5' end of pregenomic RNA and the wild type cleavage site should move the 5' end of plus DNA in the 3' direction. To this end we constructed and tested four deletion mutants: two different 4 base deletions, a precise 12 base deletion of DR1 (Δ DR1), and an 11 base deletion of sequences immediately 3' of DR1 (Figure 5A). Southern blot analysis of DNAs produced by these mutant viruses

Fig. 5. Deletions confirm 5' measurement model. (A) Sequence of the 5' end of pregenome of wild type and deletion mutants. DR1 is indicated as bold-faced sequence within box. Names of deletion clones are indicated on left. Deleted bases are indicated by -. Note in mutant Δ 4b there is also a three base substitution following the deleted region. Positions of RNA cleavages are indicated by arrows underneath respective sequences. (B) Southern blot analysis of deletion mutants. On left are indicated the positions of relaxed circular (RC), duplex linear (L), and single stranded minus (SS) DNA. Lane 1, Δ 4a viral DNA; lane 2, Δ 4b viral DNA; lane 3, Δ DR1 viral DNA; lane 4, Δ 11 viral DNA. (C) Primer extension analysis of deletion mutant DNA with oligonucleotide A. Lane 5, Δ 11; lane 6, Δ DR1; lane 7, Δ 4b; lane 8, Δ 4a; lanes 9–12, sequencing ladder generated with end labeled oligonucleotide A on cloned wild type DHBV DNA. Position of DR1 in wild type is indicated on right by vertical black bar.

indicates that mainly linear duplex DNA is present with little to no relaxed circular DNA present. Again, since normal levels of total duplex DNA are made, primer cleavage must proceed with nearly normal efficiency. Primer extension analysis of these linear duplex genomes indicates that the 5' ends of plus DNA moved towards the 3' end as predicted, though the extent of the movement was slightly imprecise. The two, four base deletions move the cleavage sites three bases, while the 12 and 11 base deletions move them 11 and 10 bases, respectively. The source of this minor (one nucleotide) discrepancy is unclear. Taken as a whole these mutants argue strongly that the site of cleavage to generate the mature primer for plus strand synthesis is a fixed distance from the 5' end of pregenomic RNA. The one nucleotide discrepancies in the deletion mutants argue that measurement from the 5' end is perhaps not the sole determinant

Fig. 6. A model for the mechanism of cleavage specificity. **Top:** As the viral RT synthesizes minus DNA (heavy arrow, with direction of synthesis indicated by arrowhead), the RNase H moiety cleaves the already copied RNA template. The processivity of RNase H action mimics the processivity of RT activity because these two moieties are covalently linked. **Bottom:** Upon completion of minus strand DNA synthesis the active site of the RT is positioned at the end of the newly completed DNA strand. This in turn positions the active site of the RNase H moiety over the 3' end of DR1, where the final cleavages take place resulting in the mature plus strand primer.

of cleavage; other, as yet undefined factors may be playing a minor role.

Two additional observations can be made about this data set. First, the Δ DR1 mutant confirms and extends our previous contention that the sequence of DR1 does not play a role in the cleavage site selection. Secondly, the fact that both insertion and deletion mutants shift the cleavage sites into different sequences again affirms that there is little sequence specificity to the cleavage reaction.

The first five nucleotides of pregenome do not direct the cleavage specificity

The experimental evidence presented thus far argues that the determinant of cleavage specificity is either a primary sequence or a structural feature found at or near the 5' end of pregenome. If it is a sequence, then this sequence must lie upstream of the 5' most insertion mutation which moved the cleavage site (the +1 and +2a mutants; Figure 4). This would implicate the first five bases of pregenome, i.e. the sequence AAGAA or any subset of this pentanucleotide. To test the model that the specificity of cleavage is determined by measurement from the sequence at the 5' end of pregenome we changed the wild type sequence, AAGAAT, to AGGCCT and GACTTA. Primer extension analysis of plus strand DNAs generated by these mutant genomes indicated that the position of cleavage was not influenced by these substitutions (data not shown). From this we conclude that the specificity of cleavage is not determined by any primary sequence, but more likely by a structural feature of the 5' end of pregenome (or the 3' end of minus DNA; see Discussion).

Discussion

We have examined the specificity of the cleavage reaction leading to the generation of the RNA primer for plus strand DHBV DNA synthesis by introducing mutations in and

around the 5' copy of DR1 in the pregenomic RNA. Our results show that the cleavage sites always remain a fixed number of nucleotides ($15-18 \pm 1$) from the 5' end of pregenome, regardless of insertions, substitutions or deletions in this region. These observations are most compatible with a mechanism involving recognition of a determinant 5' of the cleavage site, followed by length measurement and cleavage. They indicate that the cleavages are largely independent of sequence specificity, and also explain why a mechanism has evolved to exclude the so called 'precore' messages from encapsidation (Enders *et al.*, 1987; Nassal *et al.*, 1990). These mRNAs contain 5' extensions of pregenomic RNA and encode secreted versions of the viral core protein. If the precore messages were encapsidated, then upon reverse transcription the plus strand primers generated, like some of our deletion mutant primers, would be completely unrelated in sequence to DR2. Such primers are incompetent for translocation but can be extended *in situ*, generating duplex linear genomes that would be noninfectious. Our results also explain why in a prior study (Junker-Niepmann *et al.*, 1990) the addition of 15 exogenous nucleotides to the 5' end of HBV pregenomic RNA led to the production of only linear duplex viral DNA.

What is the nature of the 5' determinant? We favor a model in which the determinant for recognition is a structural feature such as the 5' cap structure of the RNA template or (more likely; see below) the 3' end of the newly synthesized minus strand DNA. How is measurement from such a structure accomplished? Clues to this may emerge from comparison with retroviral plus strand priming. On the surface these two reactions are quite distinct: retroviral plus strand primers are uncapped, derive from an internal position on the RNA and are not translocated prior to elongation. They are known to be generated by the action of the viral RNase H on DNA-RNA hybrids resulting from minus strand synthesis. These cleavages produce a 12 nucleotide polypurine tract basepaired to newly made DNA just upstream of the U3 region; this short RNA then serves as the primer. Extensive studies employing purified retroviral RT on model substrates have shed light on this reaction mechanism. Retroviral RT and RNase H activities are encoded by different regions of the *pol* gene and in some retroviruses (notably MuLV) these activities can be separated by mutation (Tanese and Goff, 1988). In fact, the MuLV RT and RNase H domains can be separately expressed in active forms as bacterial fusion proteins. While the active sites of RT and RNase H are presumably distinct in all enzymes, for other retroviruses, notably HIV, they are not always separable biochemically or genetically (Hizi *et al.*, 1990; Hostomsky *et al.*, 1991). As the polymerase activity generates nascent DNA, the RNase H activity degrades the RNA template to a series of small oligonucleotides.

Recently Oyama *et al.* (1989) have examined the products of RNase H activity generated by avian RT/RNase H during elongation of a model substrate. A synthetic DNA oligomer was used to prime synthesis off a 5' end-labeled RNA of known sequence, in the presence or absence of different ddNTPs; this allowed the synthesis of elongation products of discrete lengths. After the reaction, the labeled RNAs were recovered and sized; as expected, the length of the RNA decreased as the length of the cDNA increased. But in all cases, the lengths of the labeled RNA products indicated that 7–14 nucleotides of RNA remained annealed

to the newly-synthesized DNA. When full elongation of the minus strand DNA was allowed (i.e. in the absence of any ddNTPs), 5' oligoribonucleotides of 7–16 bases remained base-paired to the 3' end of the cDNA. Analogous experiments with HIV polymerase also revealed a similar proclivity for the RNase H cleavages to fall a fixed distance from the 3' end of the DNA (in this case 15–16 nucleotides; Furfine and Reardon, 1991). The fact that in all cases cleavage leaves behind an RNA–DNA hybrid of characteristic length has led to the suggestion that this distance may correspond to the distance between the RT and RNase H active sites on the polymerase (Oyama *et al.*, 1989; Schatz *et al.*, 1990; Furfine and Reardon, 1991).

A similar proposal would rationalize all of our findings on the generation of hepadnaviral plus strand primers (Figure 6). Like the retroviral enzyme, the hepadnaviral polymerase contains distinct homologies to known RNA-dependent polymerases and RNase H activities (Chang *et al.*, 1990; Radziwill *et al.*, 1990). Mutational analysis of the HBV P gene is consistent with the notion that, like retroviral polymerase, the hepadnaviral enzyme is organized into distinct polymerase and RNase H domains (Radziwill *et al.*, 1990). If the active sites of these domains were spaced by a distance (~ 40 Å) corresponding to ~ 15 – 18 nucleotides on their nucleic acid substrate, then upon completion of minus strand DNA synthesis the RNase H active site would be positioned over the 3' end of DR1, and cleavages there would generate capped 15–18 nucleotide primers (see Figure 6). This proposed mechanism predicts the sequence independence that we observe, and indicates that the 'measurement function' does not involve recognizing a structure on the RNA but simply polymerizing to the end of the template; the 3' end of the minus strand DNA thus positions the cleavage sites.

A sequence-independent mechanism of this type, if generally applicable, could also resolve an important remaining problem in the reverse transcription of other retroviral elements that lack LTRs (e.g. vertebrate LINE and SINE elements). For none of these elements have primer molecules been identified for second-strand synthesis. This has led some to propose that free minus strand DNA might be ligated directly into cleaved chromosomal DNA, allowing a host DNA 3' OH to serve as the plus strand primer (Voliva *et al.*, 1984). Our observations raise a more plausible alternative: that short RNA primers are generated by RNase H when the polymerase reaches the 5' end of its RNA template. Such a mechanism is particularly attractive for LINE elements, many copies of which are found to have 5' truncations attributed to premature termination of reverse transcription. *In situ* primer generation of the type described here would allow efficient synthesis of duplex linear versions of such abbreviated reverse transcripts, irrespective of where on the RNA sequence minus strand DNA synthesis terminated. Integration of such DNAs into the chromosome would then generate many of the observed genomic LINE structures.

It is instructive in this context to return to the case of retroviral plus strand priming. This reaction requires that RNase H generate an internal primer at the position of a polypurine tract. Unlike the hepadnaviral case the spacing of RT and RNase H active sites appears insufficient to account for primer generation in these viruses. We suggest that the cleavage mechanism described here is a widely

conserved feature that all RT/RNase H members, but that in some cases additional mechanisms must have evolved to account for the special feature of internal primer generation.

Materials and methods

Construction of mutant DHBV genomes

All experiments were carried out on the European strain (DHBV 3) of DHBV (Sprenkel *et al.*, 1985). Mutations were introduced into an M13 clone containing the 1363 nucleotide *Bam*HI–*Eco*RI fragment (nt 1658–3020), using the oligonucleotide directed mutagenesis procedure of Kunkel *et al.* (1987). After identification of the desired mutant by DNA sequencing, either a *Bal*I–*Nsi*I (nt 2370–2845) or *Ava*I–*Acc*I (nt 2410–2577) was transferred into a high copy number plasmid containing either the *Kpn*I–*Eco*RI (nt 1290–3020) fragment or *Bam*HI–*Eco*RI (nt 1658–3020). We referred to these clones as '1/2mers' since they contain $\sim 50\%$ of the DHBV genome. The transferred fragment was completely sequenced to circumvent the introduction of a second unwanted mutation. An overlength DNA clone, competent to support viral replication upon transfection into LMH cells, was constructed by then cloning an *Eco*RI monomer into the unique *Eco*RI site of the '1/2mer' plasmid. The consequence of this strategy is that mutations will only be found in the 5' redundancy of the subsequent RNA pregenome. These overlength clones were then used to transfect LMH cells. Clones $\Delta 4b$ and $\Delta 11$ (Figure 5) were made as described by Hirsch *et al.* (1991) and referred as $\Delta 1$ and $\Delta 2$, respectively, in that report.

Cell culture and transfections

LMH cells were maintained as described by Staprans *et al.* (1991). In each transfection experiment, the transfection efficiency was determined by cotransfection with pON249 (Geballe *et al.*, 1986), which contains the *lacZ* gene under the control of the CMV immediate early promoter. Transfections were carried out using the calcium phosphate protocol of Chen and Okayama (1987) using 12 μ g of DNA per 100 mm plate (7.5 μ g DHBV DNA + 2 μ g pON249 DNA + 2.5 μ g carrier DNA). β -galactosidase assays to determine relative transfection efficiencies were done on a portion of the sample 2–3 days post-transfection (Sambrook *et al.*, 1989).

Nucleic acid manipulations

Isolation of viral DNA from subviral core particles was usually performed 3 days post transfection as described by Staprans *et al.* (1991). Southern blot analysis was performed as described by Staprans *et al.* (1991) except sample loads were normalized for transfection efficiencies. The probe for all Southern blots was a randomly radiolabeled DHBV genome. Primer extension analyses were done as described except that up to 10 ng of core DNA was used. The oligonucleotides used in these analyses were complementary to the following positions on the DHBV genome (with reference to the *Eco*RI site as nucleotide 1): oligonucleotide A, nucleotides 2622–2599 and oligonucleotide B, nucleotides 2513–2532. The primer extension products were analyzed as described by Staprans *et al.* (1991) except when using oligonucleotide B an 8% acrylamide–8 M urea–1 $\times$ TBE gel was used during electrophoresis.

Acknowledgements

We thank Harold Varmus for thoughtful comments on the manuscript, and John Taylor for helpful and provocative discussions about retroviral RNase H. This work was supported by grants from the National Institutes of Health D.L. is supported by NIH training grant T32-HD07397-14.

References

- Baltimore, D. (1970) *Nature*, **226**, 1029–1211.
- Bartenschlager, R. and Schaller, H. (1988) *EMBO J.*, **7**, 4185–4192.
- Boeke, J.D., Garfinkel, D.J., Styles, C.A. and Fink, G.R. (1985) *Cell*, **40**, 491–500.
- Buscher, M., Reiser, W., Will, H. and Schaller, H. (1985) *Cell*, **40**, 717–724.
- Chang, L.-J., Hirsch, R.C., Ganem, D. and Varmus, H.E. (1990) *J. Virol.*, **64**, 5553–5558.
- Chen, C. and Okayama, H. (1987) *Mol. Cell. Biol.*, **7**, 2745–2752.
- Condreay, L.D., Aldrich, C.E., Coates, L., Mason, W.S. and Wu, T.-T. (1990) *J. Virol.*, **64**, 3249–3258.
- Deininger, P.L. (1989) In Berg, D.E. and Howe, M.M. (eds), *Mobile DNA*. American Society for Microbiology, Washington, DC, pp. 619–636.
- Enders, G.H., Ganem, D. and Varmus, H.E. (1985) *Cell*, **42**, 297–308.

- Finston,W. and Champoux,J.J. (1984) *J. Virol.*, **51**, 26–33.
- Furfine,E.S. and Reardon,J.E. (1991) *J. Biol. Chem.*, **266**, 406–412.
- Ganem,D. and Varmus,H.E. (1987) *Annu. Rev. Biochem.*, **56**, 651–694.
- Geballe,A.P., Spaete,R.R. and Mocarski,E.S. (1986) *Cell*, **46**, 865–872.
- Hirsch,R.C., Lavine,J., Chang,L.-C., Varmus,H. and Ganem,D. (1990) *Nature*, **344**, 552–555.
- Hirsch,R.C., Loeb,D.D., Pollack,J.R. and Ganem,D. (1991) *J. Virol.*, **65**, 3309–3316.
- Hizi,A., Hughes,S.H. and Shaharabany,M. (1990) *Virology*, **175**, 575–580.
- Hostomsky,Z., Hostomska,Z., Hudson,G.O., Moomaw,E.W. and Nodes,B.R. (1991) *Proc. Natl. Acad. Sci. USA*, **88**, 1148–1152.
- Hutchinson,C.A.III, Hardies,S.C., Loeb,D.D., Shehee,W.R. and Edgell,M.H. (1989) In Berg,D.E. and Howe,M.M. (eds), *Mobile DNA*. American Society for Microbiology, Washington, DC, pp. 593–617.
- Inuoye,S., Hsu,M.-Y., Eagle,S. and Inuoye,M. (1989) *Cell*, **56**, 709–717.
- Johnson,M.S., McClure,M.A., Feng,D.-F., Gray,J. and Doolittle,R.F. (1986) *Proc. Natl. Acad. Sci. USA*, **83**, 7648–7642.
- Junker-Niepmann,M., Bartenschlager,R. and Schaller,H. (1990) *EMBO J.*, **9**, 3389–3396.
- Kuiper,M.T.R. and Lambowitz,A.M. (1988) *Cell*, **55**, 693–704.
- Kunkel,T.A., Roberts,J.D. and Zabor,R.A. (1987) *Methods Enzymol.*, **154**, 376–382.
- Lien,J.M., Aldrich,C.E. and Mason,W.S. (1986) *J. Virol.*, **57**, 229–236.
- Lien,J.M., Petcu,D.J., Aldrich,C.E. and Mason,W.S. (1987) *J. Virol.*, **61**, 3832–3840.
- Molnar-Kimber,K.L., Summers,J. and Mason,W.S. (1984) *J. Virol.*, **51**, 181–191.
- Nassal,M., Junker-Niepmann,M. and Schaller,H. (1990) *Cell*, **63**, 1357–1363.
- Oyama,F., Kikuchi,R., Crouch,R.J. and Uchida,T. (1989) *J. Biol. Chem.*, **264**, 18808–18817.
- Pfeiffer,P. and Hohn,T. (1983) *Cell*, **33**, 781–789.
- Resnick,R., Omer,C.A. and Faras,A.J. (1984) *J. Virol.*, **51**, 813–821.
- Radziwill,G., Zentgraf,H., Schaller,H. and Bosch,V. (1988) *Virology*, **163**, 123–132.
- Radziwill,G., Tucker,W. and Schaller,H. (1990) *J. Virol.*, **64**, 613–620.
- Sambrook,J., Fritsch,E.F. and Maniatis,T. (1989) *Molecular Cloning: A Laboratory Manual*. 2nd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
- Seeger,C. and Maragos,J. (1989) *J. Virol.*, **63**, 1907–1915.
- Seeger,C. and Maragos,J. (1990) *J. Virol.*, **64**, 16–23.
- Schatz,O., Mous,J. and Le Grice,S.F.J. (1990) *EMBO J.*, **9**, 1171–1176.
- Smith,J.K., Cywinski,A. and Taylor,J.M. (1984) *J. Virol.*, **52**, 314–319.
- Sprenkel,R., Kuhn,C., Will,H. and Schaller,H. (1985) *J. Med. Virol.*, **15**, 323–333.
- Staprans,S., Loeb,D.D. and Ganem,D. (1991) *J. Virol.*, **65**, 1255–1262.
- Summers,J. and Mason,W.S. (1982) *Cell*, **29**, 403–415.
- Tanese,N. and Goff,S.P. (1988) *Proc. Natl. Acad. Sci. USA*, **85**, 1777–1781.
- Toh,H., Hayashida,H. and Miyata,T. (1983) *Nature*, **305**, 827–829.
- Temin,H. and Mizutani,S. (1970) *Nature*, **226**, 1211–1213.
- Varmus,H. and Brown,P. (1989) In Berg,D.E. and Howe,M.M. (eds), *Mobile DNA*. American Society for Microbiology, Washington, DC, pp. 53–108.
- Voliva,C.F., Martin,S.L., Hutchison,C.A.III and Edgell,M.H. (1984) *J. Mol. Biol.*, **178**, 795–813.

Received on May 17, 1991; revised on July 19, 1991